

TITLE

## PLANT VITAMIN E BIOSYNTHETIC ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/110,781, filed December 3, 1998.

5

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding vitamin E biosynthetic enzymes in plants and seeds.

BACKGROUND OF THE INVENTION

10

Vitamin E (alpha-tocopherol) is an essential element in the mammalian diet since mammals can not synthesize plastoquinones or tocopherols. The first step in the formation of plastoquinones and tocopherols in plants is the formation of homogениstate from 4-hydroxyphenylpyruvate, a reaction catalyzed by 4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27). Genetic mutants of this enzyme in *Arabidopsis* are deficient in both  
15 carotenoids and tocopherols (Norris et al. (1995) *Plant Cell* 7:2139-2149). Carotenoids (xanthophylls) in corn endosperm are valuable coloring agents in poultry feeds; tocopherols are antioxidants which may be important in oil stability and human health. Generally present at low levels in plant tissues, 4-hydroxyphenylpyruvate dioxygenase catalyzes a complex, irreversible reaction. Only recently has 4-hydroxyphenylpyruvate dioxygenase  
20 been purified to homogeneity from a plant source (Garcia et al. (1997 *Biochem. J.* 325:761-769). In plants, cDNAs encoding 4-hydroxyphenylpyruvate dioxygenase have been identified in carrots, barley and *Arabidopsis thaliana* with at least two different variants existing in this last plant. 4-Hydroxyphenylpyruvate dioxygenase is a known herbicide target (Mayonado et al. (1989) *Pestic. Biochem. Physiol.* 35:138-145; Schultz et al. (1993)  
25 *FEBS lett.* 318:162-166; Secor (1994) *Plant Phys.* 106:1429-1433). Even though plastoquinones and tocopherols are not synthesized by mammals and bacteria 4-hydroxyphenylpyruvate dioxygenase activity is present, often at high levels, and involved in phenylalanine and tyrosine degradation. Among others, the cDNAs encoding 4-hydroxyphenylpyruvate dioxygenase have been identified in *Mycosphaerella graminicola*,  
30 mice and *Coccidioides immitis*.

Plants synthesize alpha, beta, gamma and delta tocopherols of which alpha tocopherol (vitamin E) has the highest value for human nutrition. In soybean 7% of the tocopherols are vitamin E. Gamma tocopherol methyltransferase catalyzes the final step in vitamin E  
35 synthesis and has been purified to homogeneity from pepper, marigold, *Euglena* and spinach. The gene encoding gamma tocopherol methyltransferase from *Synechocystis* was identified by mutating an open reading frame encoding a methyltransferase located in the operon containing the 4-hydroxyphenylpyruvate dioxygenase gene in the *Synechocystis*

genomic sequence (Shintani and Della Penna (1998) Abstract for the American Society of Plant Physiologists meeting in Madison, WI).

Since mammals can not synthesize tocopherols, the enzymes described here may be used for the discovery of new herbicides.

5

### SUMMARY OF THE INVENTION

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 160 amino acids that has at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a corn gamma-tocopherol methyltransferase polypeptide of SEQ ID  
10 NOs:2 and 22, a rice gamma-tocopherol methyltransferase polypeptide of SEQ ID NOs:4, 6, 24, and 26, a soybean gamma-tocopherol methyltransferase polypeptide of SEQ ID NOs:8 and 28, and a wheat gamma-tocopherol methyltransferase polypeptide of SEQ ID NOs:10 and 30. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

15

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide of at least 150 amino acids that has at least 95% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of a catalpa 4-hydroxyphenylpyruvate dioxygenase polypeptide of SEQ ID  
20 NO:32, a rice 4-hydroxyphenylpyruvate dioxygenase polypeptide of SEQ ID NOs:12, 14, and 34, a soybean 4-hydroxyphenylpyruvate dioxygenase polypeptide of SEQ ID NOs:16 and 36, a Vernonia 4-hydroxyphenylpyruvate dioxygenase of SEQ ID NO:18, and a wheat 4-hydroxyphenylpyruvate dioxygenase polypeptide of SEQ ID NOs:20 and 38. The present invention also relates to an isolated polynucleotide comprising the complement of the nucleotide sequences described above.

25

<sup>sup</sup><sub>C1</sub> It is preferred that the isolated polynucleotides of the claimed invention consist of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, and 37 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, and 38. The present invention also relates to an isolated polynucleotide  
30 comprising a nucleotide sequences of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, and the complement of such nucleotide sequences.

35

The present invention relates to a chimeric gene comprising an isolated polynucleotide of the present invention operably linked to suitable regulatory sequences.

The present invention relates to an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as a yeast or a plant cell, or prokaryotic, such as a bacterial cell. The

present invention also relates to a virus, preferably a baculovirus, comprising an isolated polynucleotide of the present invention or a chimeric gene of the present invention.

The present invention relates to a process for producing an isolated host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention, the process comprising either transforming or transfecting an isolated compatible host cell with a chimeric gene or isolated polynucleotide of the present invention.

The present invention relates to a gamma-tocopherol methyltransferase polypeptide of at least 160 amino acids comprising at least 80% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 22, 24, 26, 28, and 30.

The present invention relates to a 4-hydroxyphenylpyruvate dioxygenase polypeptide of at least 150 amino acids comprising at least 95% homology based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:12, 14, 16, 18, 20, 32, 34, 36, and 38.

The present invention relates to a method of selecting an isolated polynucleotide that affects the level of expression of a gamma-tocopherol methyltransferase or a 4-hydroxyphenylpyruvate dioxygenase polypeptide in a host cell, preferably a plant cell, the method comprising the steps of: (a) constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; (b) introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; (c) measuring the level a gamma-tocopherol methyltransferase or a 4-hydroxyphenylpyruvate dioxygenase polypeptide in the host cell containing the isolated polynucleotide; and (d) comparing the level of a gamma-tocopherol methyltransferase or a 4-hydroxyphenylpyruvate dioxygenase polypeptide in the host cell containing the isolated polynucleotide with the level of a gamma-tocopherol methyltransferase or a 4-hydroxyphenylpyruvate dioxygenase polypeptide in the host cell that does not contain the isolated polynucleotide.

~~Sub C2~~ The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a gamma-tocopherol methyltransferase or a 4-hydroxyphenylpyruvate dioxygenase polypeptide gene, preferably a plant gamma-tocopherol methyltransferase or 4-hydroxyphenylpyruvate dioxygenase polypeptide gene, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a gamma-tocopherol methyltransferase or a 4-hydroxyphenylpyruvate dioxygenase amino acid sequence.

The present invention also relates to a method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a 4-hydroxy-phenylpyruvate dioxygenase or a gamma tocopherol methyltransferase polypeptide comprising the steps of: probing a cDNA or genomic library with an isolated polynucleotide of the present invention; identifying a DNA clone that hybridizes with an isolated polynucleotide of the present invention; isolating the identified DNA clone; and sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of a 4-hydroxyphenylpyruvate dioxygenase or a gamma tocopherol methyltransferase, the method comprising the steps of:

- (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a 4-hydroxyphenylpyruvate dioxygenase or a gamma tocopherol methyltransferase, operably linked to suitable regulatory sequences;
- (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of 4-hydroxyphenylpyruvate dioxygenase or gamma tocopherol methyltransferase in the transformed host cell;
- (c) optionally purifying the 4-hydroxyphenylpyruvate dioxygenase or the gamma tocopherol methyltransferase expressed by the transformed host cell;
- (d) treating the 4-hydroxyphenylpyruvate dioxygenase or the gamma tocopherol methyltransferase with a compound to be tested; and
- (e) comparing the activity of the 4-hydroxyphenylpyruvate dioxygenase or the gamma tocopherol methyltransferase that has been treated with a test compound to the activity of an untreated 4-hydroxyphenylpyruvate dioxygenase or gamma tocopherol methyltransferase, thereby selecting compounds with potential for inhibitory activity.

The present invention relates to a composition, such as a hybridization mixture, comprising an isolated polynucleotide of the present invention.

~~507~~ The present invention relates to an isolated polynucleotide of the present invention comprising at least one of 30 contiguous nucleotides derived from a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, and 37.

The present invention relates to an expression cassette comprising an isolated polynucleotide of the present invention operably linked to a promoter.

The present invention relates to a method for positive selection of a transformed cell comprising: (a) transforming a host cell with the chimeric gene of the present invention or an expression cassette of the present invention; and (b) growing the transformed host cell, preferably plant cell, such as a monocot or a dicot, under conditions which allow expression of the gamma-tocopherol methyltransferase or a 4-hydroxyphenylpyruvate dioxygenase polynucleotide in an amount sufficient to complement a null mutant to provide a positive selection means.

BRIEF DESCRIPTION OF THE  
DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

5        Figure 1 shows a comparison of the amino acid sequences of the gamma-tocopherol methyltransferase from soybean contig assembled from clones sah1c.pk001.k8, sgc5c.pk001.m23, and sah1c.pk004.g2 (SEQ ID NO:8), corn contig assembled from clones cr1n.pk0179.f10:fis, cs1.pk0065.f2, ctal1n.pk0031.d2, p0060.coran49r, and p0103.ciaay86r (SEQ ID NO:22, the 3'-end sequence of rice clone rls72.pk0010.g3 (SEQ ID NO:24), the 5'-  
10    end sequence of clone rls72.pk0010.g3 (SEQ ID NO:26), soybean clone sah1c.pk004.g2 (SEQ ID NO:28), wheat clone wr1.pk0077.f1:fis (SEQ ID NO:30), *Synechocystis sp.* (NCBI General Identifier No. 1001725; SEQ ID NO:39), and *Arabidopsis thaliana* (NCBI General Identifier No. 4106538; SEQ ID NO:40). Dashes are used by the program to maximize the alignment.

15        Figure 2 shows a comparison of the amino acid sequences of the 4-hydroxyphenylpyruvate dioxygenase from catalpa clone ncs.pk0012.g1:fis (SEQ ID NO:32), soybean clone sgc5c.pk001.j9:fis (SEQ ID NO:36), wheat clone wdk4c.pk006.m9:fis (SEQ ID NO:38), *Hordeum vulgare* having NCBI General Identifier No. 3334222 (SEQ ID NO:41), *Daucus carota* having NCBI General Identifier No. 3334219 (SEQ ID NO:42), and *Arabidopsis*  
20    *thaliana* having NCBI General Identifier No. 3334223 (SEQ ID NO:43). Dashes are used by the program to maximize the alignment.

Figure 3 depicts the distribution of the percent of alpha- and gamma-tocopherol content of 25 transgenic soybean lines transformed with the gamma tocopherol methyltransferase sequence from clone sah1c.pk001.k8.

25        Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence  
30    disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

**TABLE 1**  
**Vitamin E Biosynthetic Enzymes**

Protein	Clone Designation	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
Corn Gamma Tocopherol Methyltransferase	Contig of: cr1n.pk0179.f10 cs1.pk0065.e1 cs1.pk0065.f2 cta1n.pk0031.d2	1	2
Rice Gamma Tocopherol Methyltransferase	rl0n.pk085.e11	3	4
Rice Gamma Tocopherol Methyltransferase	Contig of: rl0n.pk099.d24 rls72.pk0010.g3	5	6
Soybean Gamma Tocopherol Methyltransferase	Contig of: sah1c.pk001.k8 sgc5c.pk001.m23 sah1c.pk004.g2	7	8
Wheat Gamma Tocopherol Methyltransferase	Contig of: wr1.pk0077.f1 wr1.pk177.b11 wle1n.pk0065.h9 wre1n.pk0111.d6	9	10
Rice 4-Hydroxyphenylpyruvate Dioxygenase	rlr12.pk0019.g5	11	12
Rice 4-Hydroxyphenylpyruvate Dioxygenase	rlr12.pk0025.e4	13	14
Soybean 4-Hydroxyphenylpyruvate Dioxygenase	Contig of: sgc5c.pk001.j9 sgs1c.pk002.a8 sfl1.pk126.n15 sgs1c.pk003.o5 sgc4c.pk001.p2	15	16
Vernonia 4-Hydroxyphenylpyruvate Dioxygenase	vs1.pk0015.b2	17	18
Wheat 4-Hydroxyphenylpyruvate Dioxygenase	Contig of: wdk4c.pk006.m9 wlmk1.pk0021.h10 wlmk1.pk0019.e2 wlm0.pk0035.d5 wlmk1.pk0013.g11	19	20
Corn Gamma Tocopherol Methyltransferase	Contig of: cr1n.pk0179.f10:fis cs1.pk0065.f2 cta1n.pk0031.d2 p0060.coran49r p0103.ciaay86r	21	22
Rice Gamma Tocopherol Methyltransferase	rls72.pk0010.g3-3'	23	24

Protein	Clone Designation	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
Rice Gamma Tocopherol Methyltransferase	rls72.pk0010.g3-5'	25	26
Soybean Gamma Tocopherol Methyltransferase	sah1c.pk004.g2	27	28
Wheat Gamma Tocopherol Methyltransferase	wr1.pk0077.f1: fis	29	30
Catalpa 4-Hydroxyphenylpyruvate Dioxygenase	ncs.pk0012.g1: fis	31	32
Rice 4-Hydroxyphenylpyruvate Dioxygenase	rlr12.pk0025.e4: fis	33	34
Soybean 4-Hydroxyphenylpyruvate Dioxygenase	sgc5c.pk001.j9: fis	35	36
Wheat 4-Hydroxyphenylpyruvate Dioxygenase	wdk4c.pk006.m9: fis	37	38

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res.* 13:3021-3030 (1985) and in the *Biochemical J.* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

#### DETAILED DESCRIPTION OF THE INVENTION

10 <sup>Sub</sup><sub>C4</sub> In the context of this disclosure, a number of terms shall be utilized. As used herein, a "polynucleotide" is a nucleotide sequence such as a nucleic acid fragment. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, 15 synthetic DNA, or mixtures thereof. An isolated polynucleotide of the present invention may include at least one of 60 contiguous nucleotides, preferably at least one of 40 contiguous nucleotides, most preferably one of at least 30 contiguous nucleotides derived from SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, or the complement of such sequences.

20 As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic

acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof.

<sup>sum</sup>  
c57 Substantially similar nucleic acid fragments may be selected by screening nucleic acid fragments representing subfragments or modifications of the nucleic acid fragments of the instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment representing at least one of 30 contiguous nucleotides derived from the instant nucleic acid fragment can be constructed and introduced into a plant or plant cell. The level of the polypeptide encoded by the unmodified nucleic acid fragment present in a plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which



result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37 and the complement of such nucleotide sequences may be used in methods of selecting an isolated polynucleotide that affects the expression of a polypeptide (vitamin E biosynthetic enzyme) in a host cell. A method of selecting an isolated polynucleotide that affects the level of expression of a polypeptide in a host cell (eukaryotic, such as plant or yeast, prokaryotic such as bacterial, or viral) may comprise the steps of: constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; measuring the level a polypeptide in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide in the host cell containing the isolated polynucleotide with the level of a polypeptide in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at

least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are at least about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above homologies but typically encode a polypeptide having at least about 50 amino acids, preferably at least about 100 amino acids, more preferably at least about 150 amino acids, still more preferably at least about 200 amino acids, and most preferably at least about 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

<sup>sub</sup>  
<sup>C4</sup> A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention

comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

“Codon degeneracy” refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

“Synthetic nucleic acid fragments” can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. “Chemically synthesized”, as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

“Gene” refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. “Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a nucleotide sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

The “translation leader sequence” refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol.* 3:225-236).

The “3' non-coding sequences” refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell* 1:671-680.

“RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary

copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into polypeptide by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a

nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several vitamin E biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other 4-hydroxyphenylpyruvate dioxygenases or gamma tocopherol methyltransferases, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In

In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

5 ~~5' to 3'~~ In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes  
 10 advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the  
 15 transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:5673-5677; Loh et al. (1989) *Science* 243:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman  
 20 and Martin (1989) *Techniques* 1:165). Consequently, a polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably one of at least 40, most preferably one of at least 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37 and the complement of such nucleotide sequences may be used in such methods to  
 25 obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide. The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a polypeptide of a gene (such as gamma tocopherol methyltransferase or 4-phenylpyruvate dioxygenase) preferably a substantial portion of a plant polypeptide of a gene, comprising the steps of: synthesizing an oligonucleotide primer  
 30 comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the  
 35 oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of a polypeptide (gamma tocopherol methyltransferase or 4-phenylpyruvate dioxygenase).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity  
5 for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol.* 36:1-34; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal  
10 or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of homogénistate in those cells. Overexpression of 4-hydroxyphenylpyruvate dioxygenase should result in a larger accumulation of homogénistate which may be used by gamma tocopherol methyltransferase to produce vitamin E. Since mammals can not synthesize tocopherols, the enzymes described herein  
15 may be used for the discovery of new herbicides.

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter  
20 sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host  
25 plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and  
30 thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different  
35 cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by directing the coding sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell* 56:247-253), signal sequences or



sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.* 100:1627-1632) with or without removing targeting sequences that are already present. While the references cited give examples of each of these, the list is not exhaustive and more  
5 targeting signals of use may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences.

10 Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

15 Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative  
20 regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages  
25 relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require  
30 the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on  
35 practical grounds, and is not an inherent part of the invention. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of

samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded vitamin E biosynthetic enzyme. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 7).

Additionally, the instant polypeptides can be used as targets to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the polypeptides described herein catalyze various steps in vitamin E biosynthesis. Accordingly, inhibition of the activity of one or more of the enzymes described herein could lead to inhibition of plant growth. Thus, the instant polypeptides could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology

outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

5 Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

10 In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; *see* Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include  
15 allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med.* 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet.* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid*  
20 *Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the  
25 instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger  
30 and Benzer (1989) *Proc. Natl. Acad. Sci USA* 86:9402-9406; Koes et al. (1995) *Proc. Natl. Acad. Sci USA* 92:8149-8153; Bensen et al. (1995) *Plant Cell* 7:75-84). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which  
35 Mutator transposons or some other mutation-causing DNA element has been introduced (*see* Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptides. Alternatively, the instant nucleic acid fragment may be used as a

hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptides can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

### EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

#### EXAMPLE 1

##### Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean, *Vernonia*, and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from cDNA Libraries from Corn, Rice, Soybean, *Vernonia*, and Wheat

Library	Tissue	Clone
cr1n	Corn Root From 7 Day Old Seedlings*	cr1n.pk0155.d1
cr1n	Corn Root From 7 Day Old Seedlings*	cr1n.pk0179.f10
cs1	Corn Leaf Sheath From 5 Week Old Plant	cs1.pk0065.e1
cs1	Corn Leaf Sheath From 5 Week Old Plant	cs1.pk0065.f2
cta1n	Corn Tassel*	cta1n.pk0031.d2
ncs	<i>Catalpa speciosa</i> Developing Seed	ncs.pk0012.g1
p0023	Corn Leaf From Plant Transformed with Gene M1C07 (leucine-rich repeat) Which Induces Resistance Prior to Genetic Lesion Formation. Harvested About One Month After Planting in Green House*	p0060.coran49r
p0103	Corn Tassel Shoots(0.1-1.4 cm)*	p0103.ciaay86r
rl0n	Rice 15 Day Old Leaf*	rl0n.pk085.e11
rl0n	Rice 15 Day Old Leaf*	rl0n.pk099.d24
rlr12	Rice Leaf 15 Days After Germination, 12 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-62 (AVR2-YAMO); Resistant	rlr12.pk0019.g5

Library	Tissue	Clone
rlr12	Rice Leaf 15 Days After Germination, 12 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-62 (AVR2-YAMO); Resistant	rlr12.pk0025.e4
rls72	Rice Leaf 15 Days After Germination, 72 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-67 (AVR2-YAMO); Susceptible	rls72.pk0010.g3
sah1c	Soybean Sprayed With Authority Herbicide	sah1c.pk001.k8
sah1c	Soybean Sprayed With Authority Herbicide	sah1c.pk004.g2
sfl1	Soybean Immature Flower	sfl1.pk126.n15
sgc4c	Soybean Cotyledon 14-21 Days After Germination (1/4 yellow)	sgc4c.pk001.p2
sgc5c	Soybean (Cotyledon 15-24 Days After Germination (3/4 yellow)	sgc5c.pk001.j9
sgc5c	Soybean (Cotyledon 15-24 Days After Germination (3/4 yellow)	sgc5c.pk001.m23
sgs1c	Soybean Seeds 4 Hours After Germination	sgs1c.pk002.a8
sgs1c	Soybean Seeds 4 Hours After Germination	sgs1c.pk003.o5
vs1	<i>Vernonia</i> Seed Stage 1	vs1.pk0015.b2
wdk4c	Wheat Developing Kernel, 21 Days After Anthesis	wdk4c.pk006.m9
wle1n	Wheat Leaf From 7 Day Old Etiolated Seedling*	wle1n.pk0065.h9
wlm0	Wheat Seedlings 0 Hour After Inoculation With <i>Erysiphe graminis f. sp tritici</i>	wlm0.pk0035.d5
wlmk1	Wheat Seedlings 1 Hour After Inoculation With <i>Erysiphe graminis f. sp tritici</i> and Treatment With Herbicide**	wlmk1.pk0013.g11
wlmk1	Wheat Seedlings 1 Hour After Inoculation With <i>Erysiphe graminis f. sp tritici</i> and Treatment With Herbicide**	wlmk1.pk0019.e2
wlmk1	Wheat Seedlings 1 Hour After Inoculation With <i>Erysiphe graminis f. sp tritici</i> and Treatment With Herbicide**	wlmk1.pk0021.h10
wr1	Wheat Root From 7 Day Old Seedling	wr1.pk0077.f1
wr1	Wheat Root From 7 Day Old Seedling	wr1.pk177.b11
wre1n	Wheat Root From 7 Day Old Etiolated Seedling*	wre1n.pk0111.d6

\* These libraries were normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.

5 \*\* Application of 6-iodo-2-propoxy-3-propyl-4(3*H*)-quinazolinone; synthesis and methods of using this compound are described in USSN 08/545,827, incorporated herein by reference.

10 cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene

Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science* 252:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

### EXAMPLE 2

#### Identification of cDNA Clones

15 <sup>sub</sup> C8 } cDNA clones encoding vitamin E biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet.* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

### EXAMPLE 3

#### Characterization of cDNA Clones Encoding

#### Gamma Tocopherol Methyltransferase

35 The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to gamma tocopherol methyltransferase from *Synechocystis* sp. (NCBI General Identifier No. 1001725). Shown in

Table 3 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), contigs assembled from two or more ESTs ("Contig"), contigs assembled from an FIS and one or more ESTs ("Contig\*"), or sequences encoding the entire protein derived from an FIS, a contig, or an FIS and PCR ("CGS"):

TABLE 3

BLAST Results for Sequences Encoding Polypeptides Homologous  
to gamma Tocopherol Methyltransferase

Clone	Status	BLAST pLog Score 1001725
Contig of: cr1n.pk0179.f10 cs1.pk0065.e1 cs1.pk0065.f2 cta1n.pk0031.d2	Contig	49.70
rl0n.pk085.e11	EST	15.10
Contig of: rl0n.pk099.d24 rls72.pk0010.g3	Contig	8.30
Contig of: sah1c.pk001.k8 sgc5c.pk001.m23 sah1c.pk004.g2:fis	CGS	64.22
Contig of: wr1.pk0077.f1 wr1.pk177.b11 wle1n.pk0065.h9 wre1n.pk0111.d6	Contig	57.00

10

Further sequencing and searching of the DuPont proprietary database allowed for the assembly of longer sequences. The BLASTX search using the nucleotide sequences from clones listed in Table 4 revealed similarity of the polypeptides encoded by the cDNAs to gamma tocopherol methyltransferase from *Arabidopsis thaliana* (NCBI General Identifier No. 4106538). Shown in Table 4 are the BLAST results for individual ESTs ("EST") or sequences encoding the entire protein derived from the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), contigs assembled from an FIS and one or more ESTs, or an FIS and PCR ("CGS"):

15

TABLE 4

BLAST Results for Sequences Encoding Polypeptides Homologous  
to gamma Tocopherol Methyltransferase

Clone	Status	BLAST pLog Score 4106538
Contig of: crln.pk0179.f10:fis cs1.pk0065.f2 cta1n.pk0031.d2 p0060.coran49r p0103.ciaay86r	CGS	118.00
rls72.pk0010.g3-3'	EST*	81.30
rls72.pk0010.g3-5'	EST*	11.00
sah1c.pk004.g2	CGS	125.00
wrl.pk0077.f1:fis	CGS	120.00

\*These sequences are derived from partial full-insert sequencing of clone rls72.pk0010.g3  
and correspond to 5'-end and 3'-end sequences.

Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID  
NOs:8, 22, 24, 26, 28, and 30 and the *Synechocystis sp.* and *Arabidopsis thaliana* sequences  
(SEQ ID NO:39 and SEQ ID NO:40, respectively). The data in Table 5 represents a  
calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:8,  
22, 24, 26, 28, and 30 and the *Synechocystis sp.* and *Arabidopsis thaliana* sequences (SEQ  
ID NO:39 and SEQ ID NO:40, respectively).

TABLE 5

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide  
Sequences of cDNA Clones Encoding Polypeptides Homologous  
to gamma Tocopherol Methyltransferase

SEQ ID NO.	Percent Identity to	
	1001725	4106538
8	37.9	56.9
22	42.6	58.6
24	42.8	67.6
26	23.8	33.3
28	41.3	62.6
30	43.2	59.5

Sequence alignments and percent identity calculations were performed using the  
Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.,  
Madison, WI). Multiple alignment of the sequences was performed using the Clustal  
method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default



parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones

5 encode one entire corn, two entire soybean, one entire wheat, and two fragments corresponding to an almost entire rice gamma tocopherol methyltransferase. These sequences represent the first corn, rice, soybean, and wheat sequences encoding gamma tocopherol methyltransferase.

#### EXAMPLE 4

##### Characterization of cDNA Clones Encoding 4-Hydroxyphenylpyruvate Dioxygenase

The BLASTX search using the EST sequences from clones listed in Table 6 revealed similarity of the polypeptides encoded by the cDNAs to 4-hydroxyphenylpyruvate dioxygenase from *Hordeum vulgare* (NCBI General Identifier No. 3334222), *Daucus carota*

15 (NCBI General Identifier No. 3334219) and *Arabidopsis thaliana* (NCBI General Identifier No. 3334223). Shown in Table 6 are the BLAST results for individual ESTs ("EST") or contigs assembled from two or more ESTs ("Contig"):

**TABLE 6**

BLAST Results for Sequences Encoding Polypeptides Homologous  
to 4-Hydroxyphenylpyruvate Dioxygenase

Clone	Status	BLAST pLog Score		
		3334222	3334219	3334223
rlr12.pk0019.g5	EST	26.70	17.70	18.00
rlr12.pk0025.e4	EST	58.52	50.40	49.40
Contig of:	Contig	55.00	81.30	107.00
sgc5c.pk001.j9				
sgs1c.pk002.a8				
sfl1.pk126.n15				
sgs1c.pk003.o5				
sgc4c.pk001.p2				
vs1.pk0015.b2	EST	39.40	51.70	52.05
Contig of:	Contig	176.00	102.00	97.00
wdk4c.pk006.m9				
wlmk1.pk0021.h1				
wlmk1.pk0019.e2				
wlm0.pk0035.d5				
wlmk1.pk0013.g1				

Further sequencing and searching of the DuPont proprietary database allowed the assembly of longer sequences and identification of 4-hydroxyphenylpyruvate dioxygenase in

other species. The BLASTX search using the nucleotide sequences from clones listed in Table 7 revealed similarity of the polypeptides encoded by the cDNAs to 4-hydroxyphenylpyruvate dioxygenase from *Hordeum vulgare* (NCBI General Identifier No. 3334222), *Daucus carota* (NCBI General Identifier No. 3334219) and *Arabidopsis thaliana* (NCBI General Identifier No. 3334223). Shown in Table 7 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS") or sequences encoding the entire protein derived from an FIS, or an FIS and PCR ("CGS"):

TABLE 7

BLAST Results for Sequences Encoding Polypeptides Homologous to 4-Hydroxyphenylpyruvate Dioxygenase

Clone	Status	BLAST pLog Score		
		3334222	3334219	3334223
ncs.pk0012.g1: fis	CGS	147.00	254.00	177.00
rlr12.pk0025.e4: fis	FIS	117.00	102.00	100.00
sgc5c.pk001.j9: fis	CGS	149.00	>254.00	>254.00
wdk4c.pk006.m9: fis	CGS	>254.00	155.00	151.00

Figure 2 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:32, 34, 36, and 38 and the *Hordeum vulgare*, *Daucus carota*, and *Arabidopsis thaliana* sequences (SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:43, respectively). The data in Table 8 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:18, 32, 34, 36, and 38 and the *Hordeum vulgare*, *Daucus carota*, and *Arabidopsis thaliana* sequences (SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:43, respectively)

TABLE 8

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to 4-Hydroxyphenylpyruvate Dioxygenase

SEQ ID NO.	Percent Identity to		
	3334222	3334219	3334223
18	49.4	58.2	60.0
32	58.5	73.3	67.4
34	86.4	73.3	69.9
36	49.4	58.2	60.0
38	94.2	59.9	58.1

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.,

Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3,  
 5 WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode substantial portions of rice, soybean, *Vernonia*, and wheat 4-hydroxyphenylpyruvate dioxygenase, and entire catalpa, soybean, and wheat 4-hydroxyphenylpyruvate dioxygenase. These sequences represent the first catalpa, rice, soybean, *Vernonia*, and  
 10 wheat sequences encoding 4-hydroxyphenylpyruvate dioxygenase.

### EXAMPLE 5

#### Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA  
 15 fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below.  
 20 Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas,  
 25 VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli*  
 30 XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment  
 35 encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10

to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic

5 proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable

10 marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene

15 from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride

20 (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles

25 resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

30 For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a

35 helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to

grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

#### EXAMPLE 6

##### Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the  $\beta$  subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

5 A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression  
10 cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50  $\mu$ L of a 60 mg/mL 1  $\mu$ m gold particle suspension is added (in order): 5  $\mu$ L DNA (1  $\mu$ g/ $\mu$ L), 20  $\mu$ l spermidine (0.1 M), and 50  $\mu$ L  $\text{CaCl}_2$  (2.5 M). The particle  
15 preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400  $\mu$ L 70% ethanol and resuspended in 40  $\mu$ L of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five  $\mu$ L of the DNA-coated gold particles are then loaded on each macro carrier disk.

20 Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the  
25 retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post  
30 bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or  
35 regenerated into whole plants by maturation and germination of individual somatic embryos.

### EXAMPLE 7

#### Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant

determined. One  $\mu\text{g}$  of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

### EXAMPLE 8

#### 5                    Evaluating Compounds for Their Ability to Inhibit the Activity    of Vitamin E Biosynthetic Enzymes

10                    The polypeptides described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 7, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant polypeptides may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("His)<sub>6</sub>"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

20                    Purification of the instant polypeptides, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the instant polypeptides are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, the instant polypeptides may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a (His)<sub>6</sub> peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include  $\beta$ -mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished



after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the instant polypeptides disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, assays for gamma tocopherol methyltransferase are presented by d'Harlingue and Camara (1985) *J. Biol. Chem.* 260:15200-15203. Assays for 4-hydroxyphenylpyruvate dioxygenase are presented by Norris et al. (1995) *Plant Cell* 7: 2139-2149.

#### EXAMPLE 9

##### Expression of gamma Tocopherol Methyltransferase in Soybean Somatic Embryos

The ability to change the levels of alpha- and gamma-tocopherol in plants by transforming them with sequences encoding gamma tocopherol methyltransferase was tested by preparing transgenic soybean somatic embryos and assaying the tocopherol levels. Plasmid DNA from clone sahlc.pk001.k8 was used as a template to prepare a Not I pcr fragment encoding the entire deduced open reading frame using the following pcr primers (forward primer AGC GCG GCC GCA TGG CCA CCG TGG TGA GGA TCC CAA CAA TCT CAT GCA TCC ACA TCC ACA; reverse primer AGC GCG GCC GCT TAT CTA GTG TGG AAA TAA TGA TCA). Standard pcr reactions were used (100 microliter total reaction containing 5 ng plasmid, 25 nmoles primer, 25 nmoles dNTPs, 1x cloned pfu buffer (Stratagene), 5% DMSO, 5U cloned pfu DNA polymerase (Stratagene). The recommended cycling parameters for pcr with pfu DNA polymerase were used (denature 45 s 94°C, anneal 45 s 55°C, extend 2.5 min 72°C, 25 cycles, final 10min extension at 72°C). The pcr product was purified on a 1% agarose/TAE gel (precast FMC), the ethidium bromide visualized band cut out and purified using a QIAquick gel extraction kit (Qiagen). The band prep (40 ng) was ligated into PCR Blunt (InVitrogen) according to the manufacturers recommendations, and the ligated plasmid used to transform *E. coli* DH10 cells. Kanamycin resistant colonies were grown overnight in liquid culture (LB/Kan), plasmids prepared and cut with Not I. Plasmids containing the correct insert size were selected for full insert sequence to confirm fidelity of pcr. Plasmids containing the correct insert verified by DNA sequencing were digested with Not I and ligated to Not I-digested and phosphatase-treated pKS67. The plasmid pKS67 was prepared by replacing in pRB20 (described in U.S. Patent No. 5,846,784) the 800 bp Nos 3' fragment, with the 285 bp Nos 3' fragment containing the polyadenylation signal sequence and described in Depicker et al. (1982) *J. Mol. Appl. Genet.* 1:561-573. Clones were screened for the sense and anti-sense orientation of the gamma tocopherol methyltransferase insert fragment by restriction enzyme digestion.

*Transformation of Soybean Somatic Embryo Cultures*

The following stock solutions and media were used for transformation and propagation of soybean somatic embryos:

Stock Solutions		Media
<u>MS Sulfate 100x stock</u>		<u>SB55 (per Liter)</u>
	(g/L)	
MgSO <sub>4</sub> ·7H <sub>2</sub> O	37.0	10 mL of each MS stock
MnSO <sub>4</sub> ·H <sub>2</sub> O	1.69	1 mL of B5 Vitamin stock
ZnSO <sub>4</sub> ·7H <sub>2</sub> O	0.86	0.8 g NH <sub>4</sub> NO <sub>3</sub>
CuSO <sub>4</sub> ·5H <sub>2</sub> O	0.0025	3.033 g KNO <sub>3</sub>
		1 mL 2,4-D (10 mg/mL stock)
<u>MS Halides 100x stock</u>		0.667 g asparagine
CaCl <sub>2</sub> ·2H <sub>2</sub> O	44.0	pH 5.7
KI	0.083	
CoCl <sub>2</sub> ·6H <sub>2</sub> O	0.00125	<u>SB103 (per Liter)</u>
KH <sub>2</sub> PO <sub>4</sub>	17.0	1 pk. Murashige & Skoog salt mixture*
H <sub>3</sub> BO <sub>3</sub>	0.62	60 g maltose
Na <sub>2</sub> MoO <sub>4</sub> ·2H <sub>2</sub> O	0.025	2 g gelrite
Na <sub>2</sub> EDTA	3.724	pH 5.7
FeSO <sub>4</sub> ·7H <sub>2</sub> O	2.784	
		<u>SB148 (per Liter)</u>
<u>B5 Vitamin stock</u>		1 pk. Murashige & Skoog salt mixture*
myo-inositol	100.0	60 g maltose
nicotinic acid	1.0	1 mL B5 vitamin stock
pyridoxine HCl	1.0	7 g agarose
thiamine	10.0	pH 5.7
		*(Gibco BRL)

5

Soybean embryonic suspension cultures were maintained in 35 mL liquid media (SB55) on a rotary shaker (150 rpm) at 28°C with a mix of fluorescent and incandescent lights providing a 16 h day 8 h night cycle. Cultures were subcultured every 2 to 3 weeks by inoculating approximately 35 mg of tissue into 35 mL of fresh liquid media.

10

Soybean embryonic suspension cultures were transformed with the plasmid containing the gamma tocopherol methyltransferase sequence by the method of particle gun bombardment (see Klein et al. (1987) *Nature* 327:70-73) using a DuPont Biolistic PDS1000/He instrument. Five µL of pKS93s plasmid DNA (1 g/L), 50 µL CaCl<sub>2</sub> (2.5 M), and 20 µL spermidine (0.1 M) were added to 50 µL of a 60 mg/mL 1 mm gold particle suspension. The particle preparation was agitated for 3 minutes, spun on a microfuge for

15

10 seconds and the supernate removed. The DNA-coated particles were then washed once with 400  $\mu$ L of 70% ethanol and resuspended in 40  $\mu$ L of anhydrous ethanol. The DNA/particle suspension was sonicated three times for 1 second each. Five  $\mu$ L of the DNA-coated gold particles were then loaded on each macro carrier disk.

5        Approximately 300 to 400 mg of two-week-old suspension culture was placed in an empty 60 mm X 15 mm petri dish and the residual liquid removed from the tissue using a pipette. The tissue was placed about 3.5 inches away from the retaining screen and bombarded twice. Membrane rupture pressure was set at 1100 psi and the chamber was evacuated to -28 inches of Hg. Two plates were bombarded, and following bombardment,  
10       the tissue was divided in half, placed back into liquid media, and cultured as described above.

15       Fifteen days after bombardment, the liquid media was exchanged with fresh SB55 containing 50 mg/mL hygromycin. The selective media was refreshed weekly. Six weeks after bombardment, green, transformed tissue was isolated and inoculated into flasks to generate new transformed embryonic suspension cultures.

20       Transformed embryonic clusters were removed from liquid culture media and placed on a solid agar media, SB103, containing 0.5% charcoal to begin maturation. After 1 week, embryos were transferred to SB103 media minus charcoal. After 5 weeks on SB103 media, maturing embryos were separated and placed onto SB148 media. During maturation  
25       embryos were kept at 26°C with a mix of fluorescent and incandescent lights providing a 16 h day 8 h night cycle. After 3 weeks on SB148 media, embryos were analyzed for the expression of the tocopherols. Each embryonic cluster gave rise to 5 to 20 somatic embryos.

30       Non-transformed somatic embryos were cultured by the same method as used for the transformed somatic embryos.

#### 25       *Analysis of Transformed Somatic Embryos*

35       At the end of the 6<sup>th</sup> week on SB148 medium somatic embryos were harvested from 25 independently transformed lines. Somatic embryos were collected in pools of five and weighed for fresh weight. Excess embryos were stored in 96-well plates at -80°. The pooled somatic embryos were lyophilized for 18 hours and the dry weight measured. The  
40       lyophilized somatic embryos were briefly pulverised with a hand held Potter homogeniser and then 600  $\mu$ L of heptane added and the samples incubated for 24 hours in the dark at room temperature to extract oils and tocopherols. The heptane was decanted and a further 300  $\mu$ L added to the samples. The extracts were combined and centrifuged (5 min, 12000 g). The supernatant was stored in amber hplc autosampler vials at -20°C prior to analysis.

45       HPLC analysis of the extracts was carried out using an HP1100 system (Agilent Technologies) 25  $\mu$ L of the heptane sample was applied to a Lichrosphere Si 60 column (5 micron 4 x 12.5 mm). The column was eluted with heptane/isopropanol (98:2 v/v) at a flow rate of 1 mL/min. After 6 minutes all four tocopherol isomers were eluted, as detected by

a HP1100 fluorescence detector (Excitation wavelength 295 nm, emission wavelength 330 nm). Individual tocopherol standards (Matreya) were diluted with hplc grade heptane to levels between 1 and 200 ng/ul to construct a 6 point external standard curve. Tocopherols in each oil were quantified using a standard curve run on the same day as the samples. The

5 sum of tocopherol peak areas of samples from a non-transformed control line were compared with those of 25 independent gamma tocopherol methyltransferase-transformed, hygromycin resistant lines. Figure 3 shows a graph depicting the distribution of the percent alpha- and gamma-tocopherol of soybean somatic embryos transgenic for the gamma tocopherol methyltransferase gene and a control line. The mean alpha tocopherol content is 40 to 70%.

10 Some lines, such as the ones represented in bars 2, 5, and 18 appear to represent overexpression of the gamma tocopherol methyltransferase, which would be predicted to yield higher levels of alpha tocopherol. Other lines, such as those represented in bars 6 and 10 appear to represent co-suppression of the gamma tocopherol methyltransferase since this is predicted to yield higher levels of gamma tocopherol. These results indicate that

15 transgenic expression of gamma-tocopherol methyltransferase affords the ability to manipulate tocopherol levels as desired for a particular application.

Various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

20 The disclosure of each reference set forth above is incorporated herein by reference in its entirety.